AWARD NUMBER: W81XWH-18-1-0421

TITLE: Discovery and Characterization of Functional Breast Cancer Microproteins

PRINCIPAL INVESTIGATOR: Alan Saghatelian

CONTRACTING ORGANIZATION: The Salk Institute for Biological Studies

REPORT DATE: NOVEMBER 2021

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
NOVEMBER 2021	Final	1 Aug 2018 – 31 Jul 2021

4. TITLE AND SUBTITLED

Discovery and Characterization of Functional Breast Cancer Microproteins

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-18-1-0421

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Saghatelian, Alan

E-Mail: asahatelian@salk.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Salk Institute for Biological Studies
10010 N Torrey Pines Rd
La Jolla, CA 92037-1002

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

There were two significant findings during the research period. First, in collaboration with Dr. Adolfo Ferrando we found that TINCR is a tumor suppressor in squamous cell carcinoma. Second, we completed our CRISPR/Cas9 screen and identified nearly 1000 smORFs that regulate breast cancer growth. An amazing finding that we will validate and further characterize.

15. SUBJECT TERMS
Breast cancer, disease genes, small open reading frames (smORFs), microproteins, smORF oncogene, smORF tumor suppressor, CRISPR, TINCR, NuRD complex

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRDC
U	U	U	UU	10	19b. TELEPHONE NUMBER *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

TABLE OF CONTENTS

1. **INTRODUCTION:** The goal of this proposal is to identify novel genetic drivers of breast cancer. Previous genome-wide screens for genes that drive breast cancer did not include a set of human genes that contain small open reading frames (smORFs). smORFs encode peptides and small proteins (microproteins) that are less than 100 amino acids in length. The human genome project missed smORFs because gene-finding algorithms utilized a length cutoff to try limit false positives, but in the process missed thousands of protein-coding smORFs. Our hypothesis is that there are smORFs with critical roles in breast cancer such as smORF oncogenes, tumor suppressors, and synthetic lethal genes. Here, we endeavor to identify these and characterize cancer-driving smORFs using two approaches. The first approach focuses on an RNA called TINCR that has been linked to breast cancer. TINCR is thought to be 'non-coding' but we discovered that TINCR contains a smORF that encodes an 87-amino acid microprotein. **We will test whether this microprotein is responsible for the deleterious role of TINCR reported in breast cancer.** Second, we will take an unbiased screening approach to identify smORFs with roles in breast cancer cell survival and invasion. This strategy begins by using our smORF-discovery platform to identify smORFs in HCC1954, HMEC, MCR-7, and MDA-231 cell lines (note: the MCF-7 and MDA-MB-231 lines were added based on a referee comments). **We will then perturb these smORF-encoding genes to identify those smORFs with roles in driving the phenotypes associated with breast cancer, with the goal of finding oncogenes or tumor suppressors.** Our recent discovery of thousands of novel protein-coding genes has revealed an unexpected blind spot in gene annotation methods and in this proposal we examine whether any of these new genes have roles in breast cancer.

2. **KEYWORDS:** breast cancer, disease genes, small open reading frames (smORFs), microproteins, smORF oncogene, smORF tumor suppressor, CRISPR, TINCR, NuRD complex.

3. **ACCOMPLISHMENTS:**

   - **What were the major goals of the project?**
   - Specific Aim 1. Breast cancer microproteins from smORFs on non-coding RNAs that have already been linked to breast cancer. (100% complete)
   - Specific Aim 2. Use Ribo-Seq to define the smORFeome of HCC1954 breast cancer cells and human mammary epithelial cells (HMECs).
   - UPDATED Specific Aim 2. Use Ribo-Seq to define the smORFome of HCC1954, HMEC, MCF-7, and MDA-MB-231 breast cancer cells. We now include MDA-MB-231 (as requested by a referee) and the MCF-7 cell line, which is the comparator for MDA-MB-231 line. (100% complete)
   - **What was accomplished under these goals?**

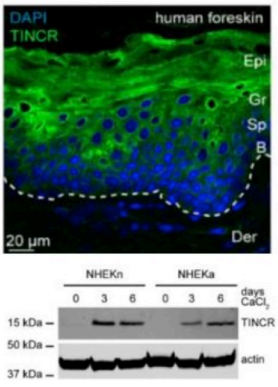

Figure 1. **Expression of TINCR protein. Top)** Immunofluorescence detection of TINCR protein in human skin. Der:dermis, B: basal layer, Sp: spinocellular layer, Gr: granulose layer, Epi: epidermis. Dashed line indicates the dermal-epidermal boundary. **Bottom)** Western blot analysis of endogenous TINCR microprotein in neonatal (NHEKn) and adult (NHEKa) normal human keratinocytes upon *in vitro* differentiation following calcium chloride treatment.

Specific Aim 1

Major task 1: Generation of Anti-TINCR-MP antibodies Specific (**100% complete**). The investigation of microproteins requires antibodies to accurately assess endogenous levels, or overexpression or knockdown of the microprotein. No anti-TINCR microprotein (anti-TINCR-MP) antibodies exist so we generated our own antibodies using methods developed over the last 40 years at the Salk institute. We raised two sets of antibodies to ensure that we would obtain antibodies of sufficient quality for biological studies. We have validated these antibodies in cell lines overexpressing TINCR.
During this work, we discovered that Professor Adolfo A. Ferrando at Columbia University was studying the role of TINCR in driving melanoma. We established a collaboration to use our antibodies to stain tissues and perform Western blots and immunoprecipitation experiments. Using this reagent, we obtained excellent staining of TINCR microprotein in tissues and changes in TINCR microprotein levels (Fig. 1). This data is part of a manuscript that has been submitted and is under review.

Major Task 2: Identify the NuRD-binding domain within TINCR-MP by testing TINCR-MP mutants for their functional impact on the

immunoprecipitation of NuRD proteins (HDAC1, RBPP7) from MDA-MB-453 cells (**100% complete**). We have prepared the TINCR-MP deletion constructs (i.e. TINCR-MP-FLAG deletion from 2-10, TINCR-MP-FLAG deletion from 11-20… TINCR-MP-FLAG deletion 80-87 (a total of 9 constructs). We have developed a robust immunoprecipitation assay for the interaction between NuRD proteins and TINCR-MP. Because TINCR is interacting with a protein complex involved in gene regulation, we have performed RNA-Seq experiments to determine if this TINCR can modulate gene expression. We find that TINCR regulates a series of genes that align with biological processes involved in wound healing and cancer (**Fig. 2**), consistent with our original hypothesis about this protein. Furthermore, these data support our proposed experiments to delineate TINCR regulated phenotypes and biology. However, immunoprecipitation experiments in MDA-MB-453 did not identify NuRD as a target complex so its role in breast cancer biology differs from that of the skin. Dr. Martinez will continue to pursue questions regarding TINCR in his newly established lab at University of California, Irvine.

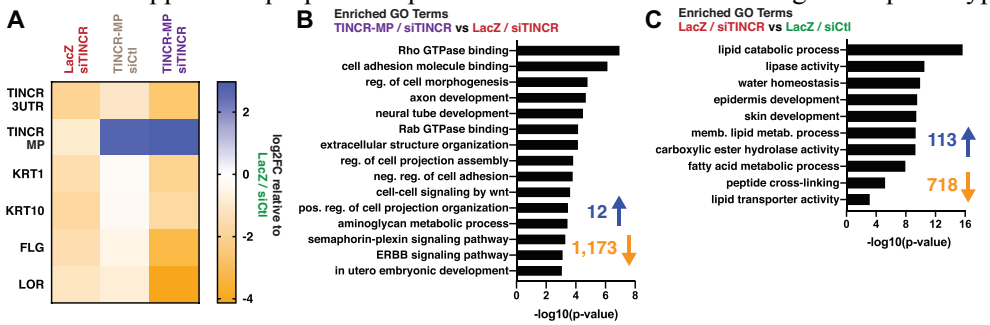

Figure 2. **(A)** Heatmap of log2 fold-changes (log2FC) in RNA expression of differentiation genes in 3D organotypic skin models relative to LacZ/siCtl control. TINCR-3UTR is representative of endogenous TINCR expression, while TINCR-MP measures change due to TINCR-MP retrovirus infection. **(B-C)** Significantly enriched gene ontology (GO) terms for TINCR-MP rescue **(B)** and TINCR-depletion **(C)** regulated genes. Number of significantly upregulated genes (blue) and downregulated genes (orange) are also shown ($|log2FC| > 1$, $p_{adj} < 0.05$). *TINCR-MP regulated genes suggest it is a regulator of wound healing processes.*

Major Task 3: Knockdown of TINCR RNA and rescue experiments with wild type TINCR-MP and TINCR-MP mutants in MDA-MB-453 cells (**100% complete**). To establish whether TINCR-MP effects cancer proliferation, we measured the viability of LNCaP cells transfected with WT TINCR-MP or Mut TINCR-MP expression constructs by WST-1 assay. TINCR-MP overexpression reduced proliferation relative to the translation impaired mutant and empty vector control transfection, consistent with the published study showing overexpression of TINCR RNA suppresses proliferation of LNCaP cells (9). In breast cancer cells, we did not observe a loss-of-function phenotype for TINCR, suggesting that its activity is due to a gain-of-function, which Dr. Martinez is pursuing in his lab.

Major Task 4: Identify microprotein-protein interactions for newly discovered smORFs from the NEAT1, LINC00346, ARRDC1-AS1, DLEU1, and LINC00958 ncRNAs (**100% complete**). We did not find any robust interacting proteins for these target genes using current platforms, but will continue to attempt using alternative methods since microprotein interactions are challenging for several reasons.

Specific Aim 2

Major Task 1: Ribo-Seq characterization of the smORFome of HCC1954 and HMECs. **Updated Major Task 1**: Ribo-Seq characterization of the smORFome of HCC1954, HMECs, MCF-7, and MDA-MB-231 cells (**100% complete**). We have completed the Ribo-Seq and RNA-Seq of the MCF-7, MDA-MB-231, and HCC1954 cells.

Major Task 2: Use CRISPR/Cas9 to screen for smORFs that mediate breast cancer cell proliferation by targeting smORFs that are specific to HCC1954 cells. **Updated Major Task 2:** Use CRISPR/Cas9 to screen for smORFs that mediate breast cancer cell proliferation by targeting smORFs that are in MDA-MB-231 and MCF-7 cells (**100% complete**).

Ribosome profiling of three breast cancer cell lines, HCC1954, MCF7 and MDA-MB-231 identified 8,262 previously unannotated smORFs (**Fig. 1**). A custom guide RNA (gRNA) library was designed using CCTop[22] targeting the complete set of breast cancer microproteins— 53,000 gRNAs with an average 6 gRNAs per smORFs. Additionally, 300 genes with known roles in housekeeping, cell cycle, apoptosis, and cell growth (4 gRNAs/gene from a known library[23]) were included as positive controls. I decided to perform the screen in MCF7 (ER+) and MDA-MB-231 (triple negative) cell lines to reveal any microproteins that might be context dependent. Stable Cas9-expressing MCF7 and MDA-MB-231 were infected with the lentiviral library with a higher ratio of cells such that a cell receives none or one gRNA (<0.3 MOI). Following the infection, the puromycin treatment for three days removed the uninfected cells and the remaining 30-40 million cells were harvested which served as the baseline. An equal number of cells were plated for culture and the cells were passaged routinely every 3 days. At days 14 and 29, cells were harvested for analysis that consisted of genomic DNA extraction, sample preparation and sequencing (**Fig. 2A**). The resultant sequencing data was analyzed with MAGeCK pipeline[24] to identify genes that dropped out over the timeline of the experiment. The results revealed hundreds of microproteins that impair the growth of breast cancer cells (**Fig. 2B-D**). The screen verified the important growth regulators such as cell cycle protein CDK1 and the apoptosis inhibitor BIRC5. KRAS was found to have a pronounced effect upon loss in MDA-MB-231 cell line when compared to MCF7, and with FOXA1 the inverse was observed, which is consistent with previous screens in these cells[25,26] (**Fig. 2B, 2C**). These data validate the screen and increase the confidence in identification of microproteins that are cell line specific. Most of the hits were identified in either MDA-MB-231 cell line or MCF7 cell lines, with 79 microproteins affecting cell-growth in both cell lines (**Fig. 2D**).

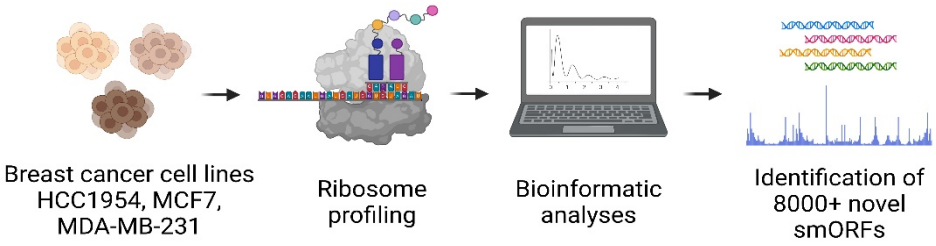

Figure 1. Discovery of microprotein encoding smORFs by ribosome profiling.

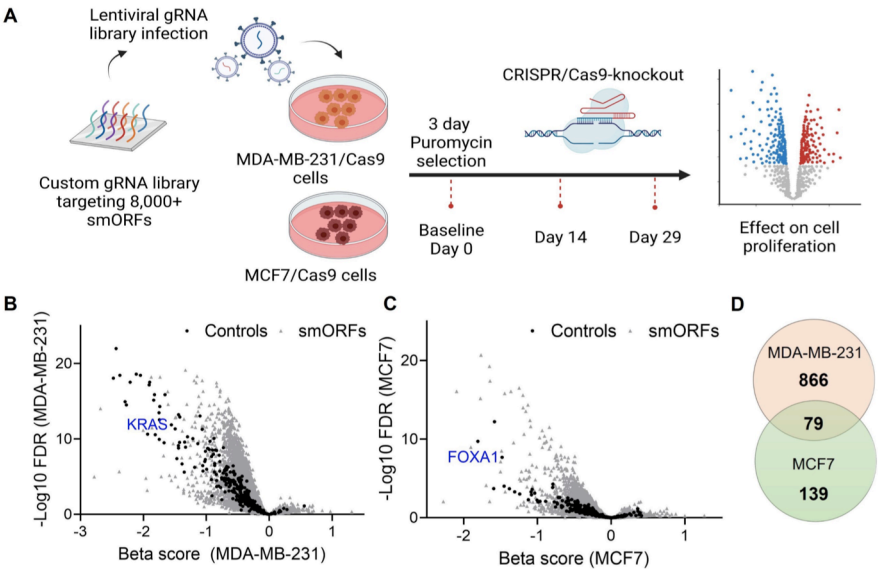

Figure 2. Experimental design of CRISPR/Cas9-drop out screen targeting smORFeome in breast cancer cell lines. Volcano plots depicting the results of the screen with one example of control gene shown in (B) MDA-MB-231 (C) MCF7 cell line. (D) The number of smORFs identified in each cell line with false discovery rate (FDR<0.005) and beta score ≤-0.5 or ≥0.5.

- **What opportunities for training and professional development has the project provided?**

Dr. Martinez obtained an NIH K01 award based on his preliminary results from TINCR and has obtained an independent academic position at University of California, Irvine as an Assistant Professor.

- **How were the results disseminated to communities of interest?**

We have a manuscript detailing our TINCR results stated above under review at JCI.

- **What do you plan to do during the next reporting period to accomplish the goals?**

This is the final report for this project period but we have applied for an expansion award to follow up on the results from the CRISPR screen.

4. IMPACT:

- **What was the impact on the development of the principal discipline(s) of the project?**

We have learned that TINCR regulates cell proliferation and is a driver of skin cancer (collaboration with Columbia team). We have also revealed the potential importance of microproteins in breast cancer through our CRISPR screen and will endeavor to show that these genes are disease genes and potential therapeutic targets.

- **What was the impact on other disciplines?**

More generally, these results will highlight smORFs/microproteins as disease regulators in biology.

- **What was the impact on technology transfer?**

Nothing to Report now, but the identification of specific microproteins in breast cancer and targeting these genes will result in new technology.

- **What was the impact on society beyond science and technology?**

These findings when disseminated can inform on the clinical treatment of breast cancer and spur new drug development programs.

5. CHANGES/PROBLEMS:

- **Changes in approach and reasons for change**

Nothing to Report

- **Actual or anticipated problems or delays and actions or plans to resolve them**

Nothing to Report

- **Changes that had a significant impact on expenditures**

Nothing to Report

- **Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents**

Nothing to Report

- **Significant changes in use or care of human subjects**

Nothing to Report

- **Significant changes in use or care of vertebrate animals.**

Nothing to Report

- **Significant changes in use of biohazards and/or select agents**

Nothing to Report

6. PRODUCTS:

- **Publications, conference papers, and presentations**

  - **Journal publications.**

Lucia Morgado-Palacin, Jessie A. Brown, Thomas Martinez, Juana M. Garcia-Pedrero, Farhad Forouhar, S. Aidan Quinn, Clara Reglero, Joan Vaughan, Sandra Rodriguez-Perales, Eva Allonca, Rocio Granda-Diaz, Agustin F. Fernandez, Mario F. Fraga, Arianna L. Kim, Jorge Santos-Juanes, David M. Owens, Juan P. Rodrigo, Alan Saghatelian, Adolfo A. Ferrando. The TINCR ubiquitin-like microprotein is a tumor suppressor in squamous cell carcinoma. Journal of Clinical Investigation.

Submitted and under review; Acknowledgment of federal support, yes.

  - **Books or other non-periodical, one-time publications.**

Nothing to Report.

  - **Other publications, conference papers, and presentations.**

Presented this work at several national meetings including the American Chemical Society National Meeting, and a Gordon Research Conference on Translation.

- **Website(s) or other Internet site(s)**

Lab website: https://saghatelian.salk.edu

- **Technologies or techniques**

The technologies we use are based on next-generation sequencing technologies and they will be disseminated in our published manuscript. In addition, we will have reagents such as antibodies and plasmids that we will share freely with any scientist that wants to use them.

- **Inventions, patent applications, and/or licenses**

Nothing to Report

- **Other Products**

Our databases of smORFs will be provided in our published manuscript and we also submit all the raw data to the appropriate national databases for others to use as they wish. As mentioned, we will also have antibodies, and DNA constructs for smORF overexpression, as well as shRNA and CRISPR libraries that we will share.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

- **What individuals have worked on the project?**

Name:	*Alan Saghatelian*
Project Role:	*PI, no change*
Researcher Identifier (e.g. ORCID ID):	0000-0002-0427-563X
Nearest person month worked:	0.5
Contribution to Project:	*Oversaw the project.*
Funding Support:	*Frederik Paulsen Chair/NIH/Clayton Foundation*

Name:	*Archita Agrawal*
Project Role:	*Postdoctoral Fellow*
Researcher Identifier (e.g. ORCID ID):	0000-0001-9448-420X
Nearest person month worked:	3
Contribution to Project:	*Dr. Agrawal performed the CRISPR screen.*
Funding Support:	*Frederik Paulsen Chair/Clayton Foundation*

Name:	*Thomas Martinez*
Project Role:	*Postdoctoral Fellow*
Researcher Identifier (e.g. ORCID ID):	0000-0002-4011-8164
Nearest person month worked:	6
Contribution to Project:	*Dr. Martinez handled all TINCR work and oversaw all data analysis for the smORF discovery.*
Funding Support:	*NIH and Ferring Foundation*

Name:	*Almudena Garcia Ruiz*
Project Role:	*Postdoctoral Fellow*
Researcher Identifier (e.g. ORCID ID):	0000-0003-4613-1818
Nearest person month worked:	3
Contribution to Project:	*Dr. Garcia Ruiz assisted with cell culture experiments and RNA-Seq data collection.*
Funding Support:	*NIH and Clayton Foundation*

Name:	*Lina Xie*
Project Role:	*Postdoctoral Fellow*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	4
Contribution to Project:	*Dr. Xie assisted with the design of guide RNA libraries.*
Funding Support:	*Clayton Foundation*

Name:	*Cynthia Donaldson*
Project Role:	*Lab Coordinator*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	2
Contribution to Project:	*Mrs. Donaldson handled all cell culture, plasmid generation, Ribo-Seq, and RNA-Seq experiments.*
Funding Support:	*NIH and Clayton Foundation*

- **Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?**

Yes.

Completed:
NIH/NIDDK 5RC2DK114785-04 (Saez PI, Scripps Research), Chemoproteomic Identification and Therapeutic Validation of Proteins of Metabolic Significance, annual direct, 08/01/2017-07/31/2021, 5% effort. The major goal of this project is to elucidate the biochemical function of the lipases Aig1 and ADTRP, and to use technology developed in the Saghatelian lab to identify protein-metabolite interactions for proteins of interest to the Saez lab with functions in metabolism and metabolic disease.

NIH/NIGMS 5R01GM102491-09 (Saghatelian PI, Salk Institute), The Discovery of Human Peptide Encoding Genes, 09/01/2017-04/30/2021, annual direct, 25% effort. The major goal of this project is to annotate and functionally characterize a large group of protein-coding genes that were missed by the human genome project--to annotate all human and mouse small protein-coding open reading frames (smORFs), and characterize the biochemical and cellular functions of selected smORFs.

NIH/NIDDK 5R01DK106210-05 (Kahn PI, Beth Israel Deaconess Medical Center), Regulation of the biosynthesis of a novel class of anti-diabetic lipids, 04/20/2016-03/31/2021, annual direct, 5% effort. The major goal of this project is to understand how PAHSAs are regulated in vivo. The Saghatelian group will perform the majority of the quantitative LC-MS experiments.

NIH/NIMH 5R01MH113905-03 (Chalasani PI, Salk Institute), Dissecting molecular elements of threat behavior, 07/07/2017-10/31/2019, annual direct, 5% effort. The major goal of this project is to understand the neural mechanisms that encode threat responses (both behavioral and physiological) in an invertebrate model system.

New:
NIH/NIDDK, 2R01DK106210-06 (Kahn PI, Beth Israel Deaconess Medical Center), Mechanisms for regulation of a novel class of anti-diabetic lipids, 04/01/2021-03/31/2025, annual direct, 10% effort. The major goal of this project is to study the role of FAHFA biosynthesis and degradation in animal models of diabetes, and to identify additional genes that regulate FAHFAs using mouse genetics.

NIH/NIDDK, 1RC2DK129961-01 (Cohen PI, Rockefeller University), An Encyclopedia of the Adipose Tissue Secretome to Identify Mediators of Health and Disease, 07/15/2021-06/30/2026, annual direct, 12.5% effort. The major goal of this project is to identify secreted factors from metabolic tissues that regulate metabolism.

- **What other organizations were involved as partners?**

Nothing to Report

8. SPECIAL REPORTING REQUIREMENTS

- **COLLABORATIVE AWARDS:** Not applicable.
- **QUAD CHARTS:** Not applicable.

9. APPENDICES: Award Chart, Award Expiration Transition Plan.